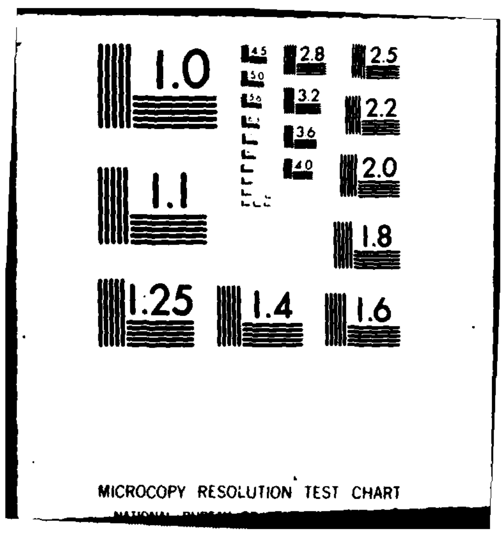
AD-A111 807 CALIFORNIA UNIV BERKELEY F/G 5/3
A BAYESIAN INTERPRETATION OF DATA TRIMMING TO REMOVE EXCESS CLA--ETC(U)
MAR 81 W S JEWELL AFOSR-81-0122

UNCLASSIFIED NL

1 of 1

END
DATE
FILMED
4-82
DTIC


1.0
4.5
5.0
5.6
2.8
2.5
3.2
2.2
3.6
4.0
2.0
1.1
1.8
1.25
1.4
1.6
MICROCOPY RESOLUTION TEST CHART

ADA111807

A BAYESIAN INTERPRETATION OF
DATA TRIMMING TO REMOVE EXCESS CLAIMS

William S. Jewell
University of California, Berkeley
& ETH-Zürich

Abstract

The effect of excess or catastrophic claims is well recognized in insurance. For example, in experience rating it is customary to truncate the data to minimize the effect of such outliers; Gisler has recently proposed a credibility formula using such data trimming. This paper develops a model of the excess claims process and finds the exact Bayesian forecast. The resulting forecast form is approximately a data trim, thus justifying the simpler, heuristic approach.

Zürich, March, 1981

AFOSR-81-0122

†Reproduction was supported by the Air Force Office of Scientific Research (AFSC), USAF, under Grant AFOSR-81-0122. Reproduction in whole or in part is permitted for any purpose of the United States Government.

1. Introduction

The effect of excess or catastrophic claims is well recognized in insurance. Typically, one wishes not only to analyze them in detail to determine and, if possible, correct their causes, but also to modify the data so as to minimize their effect upon normal operating procedures of the firm.

For example, in experience rating, data $\underline{x} = (x_1, x_2, \ldots x_n)$ collected from a policyholder's experience in years $1, 2, \ldots n$ is used to modify his premium for year $n+1$. If $\tilde{y} = \tilde{x}_{n+1}$ is the random variable denoting next year's total paid claims, the fair premium will be just the regression of $\tilde{y}$ on the data $\underline{x}$, or $\mathcal{E}(\tilde{y}|\underline{x})$. In credibility theory, it is assumed that this forecast is linear in the data, giving the well-known formula:

$$\mathcal{E}(\tilde{y}|\underline{x}) \approx f(\underline{x}) = (1-Z_n)m + Z_n\left(\frac{1}{n}\sum_{t=1}^{n} x_t\right) . \tag{1.1}$$

Here m is the "manual" (fair, no-data) premium, and $Z_n = n/(n+N)$ is the credibility factor with time constant N determined empirically or from a Bayesian model (see, e.g., Norberg(1979) for further details).

The effect of an excess claim upon experience rating is obvious from (1.1). What one would like to do is to detect and remove this claim from the data, and spread all or a portion of the excess amount over all the policyholders, perhaps by charging it against a special reserve. However, in many situations it is not possible or economical to use qualitative information about the claim to decide if it is of ordinary or excess type, and one must use a numerical procedure to "cleanse" the data before using (1.1). Based upon heuristic methods used in industry, A. Gisler (1980) proposed to replace (1.1) by:

A

$$f(\underset{\sim}{x}) = a + b\left(\frac{1}{n}\sum_{t=1}^{n}\min(x_t, M)\right) \quad , \tag{1.2}$$

where the parameters (a,b,M) are adjusted so as to minimize the mean-squared error in the forecast; the result could be called a data-trimmed credibility formula.

2. A Bayesian Model for Outliers

We now develop a model which describes how excess claims arise, and then find the exact Bayesian prediction formula. By comparing this with Gisler's form (1.2), we will be able to provide additional motivation for the trimming procedure.

First of all, we assume that an ordinary claim random variable, $\tilde{x}_o$, has a known density, $p_o(x_o|\theta)$, depending upon an unknown parameter θ which characterizes the different policyholders and their exposure characteristics. The first two moments of this random variable are:

$$m_o(\theta) = \mathcal{E}(\tilde{x}_o|\theta) \quad ; \quad v_o(\theta) = \mathcal{V}(\tilde{x}_o|\theta) \ . \tag{2.1}$$

In the usual experience rating model, we are given several independent observations of the $\tilde{x}_o$ type from a policy with fixed, but unknown, θ, and we wish to estimate the mean of the next observation from the same policy. Th' is equivalent to estimating $m_o(\theta)$, given a prior densi·y on θ, and the data $\underset{\sim}{x}$. If it is known that all data is of ordinary type, then in many cases the credibility forecast (1.1) is exact, or a good approximation.

Now, however, suppose that it is occasionally possible that we observe instead an excess claim random variable, $\tilde{x}_e$, with density $p_e(x_e)$ not depending upon θ (although this can be easily generalized, if desired). This excess

claim is considered to be the result of some extraordinary cause, so that the density p_e will have large mean and variance compared with every density p_o. We also assume that there is no qualitative way in which one can identify an excess claim as such; thus, the densities should have overlapping ranges, otherwise, there would be no difficulty in separating excess claims based upon their magnitude.

We continue to let $\underline{x} = (x_1, x_2, \ldots x_n)$ represent the observational data, assumed independent, given θ. But the <u>observation</u> random variable, $\tilde{x}_t$, $(t=1,2,\ldots n)$, is now sometimes an ordinary, sometimes an excess random variable, and we assume that there is a known <u>contamination probability</u>, π, that independently selects if an ordinary claim is replaced by an excess claim. In other words, we assume that the individual observations follow the mixed density:

$$(2.2) \qquad p(x_t|\theta) = (1-\pi)p_o(x_t|\theta) + \pi p_e(x_t) ,$$

so that the likelihood of $\underline{x}$, *

$$(2.3) \qquad p(\underline{x}|\theta) = \prod_{t=1}^{n} p(x_t|\theta) ,$$

consists of 2^n terms. Since π is small, however, only the first few terms will generally be significant (e.g., there are usually only no, one, or a few excess claims in any small sample).

As in other experience rating models, we assume that we are given a prior density, $p(\theta)$, on the unknown parameter, so that Bayes' law then gives a posterior-to-data density for the unknown parameter of:

$$(2.4) \qquad p(\theta|\underline{x}) = \frac{p(\underline{x}|\theta)\; p(\theta)}{p(\underline{x})} ,$$

---

*From this point on, we are using the usual Bayesian trick of using p(.) for several different densities, and letting the variables "speak for themselves".

where $p(\underline{x})$ is the integral of the numerator over θ. Now, however, we must remember that we are not interested in predicting just the next observation, but rather in predicting the next observation, given that it is of ordinary type; this random variable, call it $\tilde{y}_o$, has a density $p_o(y_o|\theta)$ if θ were known. It follows then that, given the data $\underline{x}$, we can form the Bayesian predictive density of $\tilde{y}_o$ from (2.4) as follows:

$$(2.5) \qquad p(y_o|\underline{x}) = \int p_o(y_o|\theta)\ p(\theta|\underline{x})\ d\theta \quad .$$

The exact Bayesian mean predictor of y_o is then just the first moment:

$$(2.6) \qquad \mathcal{E}(\tilde{y}_o|\underline{x}) = f(\underline{x}) = \int m_o(\theta)\ p(\theta|\underline{x})\ d\theta \quad .$$

To better understand how this formula depends upon the data, we need to develop further the likelihood (2.3).

3. Single Observation Case

First, suppose that only n=1 observation has been made. Then (2.3) has only two terms, and the exact forecast is:

$$(3.1) \quad f(x_1) = \frac{1}{p(x_1)}\Big[(1-\pi)\int m_o(\theta)\ p_o(x_1|\theta)\ p(\theta)\ d\theta + \pi\, p_e(x_1)\int m_o(\theta)\ p(\theta)\ d\theta\Big] \ ,$$

where

$$(3.2) \ p(x_1) = (1-\pi)p_o(x_1) + \pi p_e(x_1) \ ; \ p_o(x_1) = \int p_o(x_1|\theta)p(\theta)d\theta \ .$$

The second integral in (3.1) is just the a priori expected

value of $\tilde{y}_o$ (the manual premium):

$$(3.3) \qquad m_o = \mathcal{E} m_o(\tilde{\theta}) = \mathcal{E}\tilde{y}_o \;,$$

which would be the "forecast" if no data were available.

The first integral in (3.1) is more interesting, as it is related to the Bayesian prediction in which <u>it is known that the observation is of ordinary type</u>. In contrast to $f(x_1)$, which is the prediction from an arbitrary observation following (2.2), we can <u>define</u> $f_o(x_1)$ as the <u>ordinary observation prediction</u>, gotten from (2.6) by setting $\pi=0$:

$$(3.4) \quad \mathcal{E}(\tilde{y}_o | x_1 \text{ ordinary}) = f_o(x_1) = \int m_o(\theta) \frac{p_o(x_1|\theta)\; p(\theta)}{p_o(x_1)} d\theta \quad .$$

This could, of course, follow the linear credibility law (1.1).

Finally, we rewrite the exact forecast as:

$$(3.5) \qquad f(x_1) = \frac{1}{p(x_1)} \left[(1-\pi) p_o(x_1) f_o(x_1) + \pi p_e(x_1) m_o \right] \;,$$

which can be rewritten in two revealing forms: the first,

$$(3.6) \qquad f(x_1) = \frac{f_o(x_1) + \phi(x_1) m_o}{1 + \phi(x_1)} \;,$$

with

$$(3.7) \qquad \phi(x_1) = \left(\frac{\pi}{1-\pi}\right) \frac{p_e(x_1)}{p_o(x_1)}$$

as an "odds-likelihood-ratio"; the second in a credibility format:

$$(3.8) \qquad f(x_1) = \left[1 - Z(x_1)\right] m_o + Z(x_1) f_o(x_1) \;,$$

with a new data-dependent credibility factor:

$$Z(x_1) = \left[1+\phi(x_1)\right]^{-1} = \frac{(1-\pi)p_o(x_1)}{(1-\pi)p_o(x_1)+\pi p_e(x_1)} \quad . \tag{3.9}$$

$Z(x_1)$ is essentially the a posteriori probability that the observation x_1 is ordinary.

In the usual situation, the averaged ordinary density $p_o(x_1)$ and the excess density $p_e(x_1)$ might appear as in Figure 1, giving then the weighting functions $\phi(x_1)$ or $Z(x_1)$ shown in Figure 2.

4. Comparison with Trimming in the Credibility Case

As discussed in the first Section, it is often the case that $f_o(\underline{x})$ is linear in the data $\underline{x}$, i.e. it follows (1.1), with m replaced by m_o, and the ordinary (non-data-dependent) credibility factors Z_n replaced by:

$$Z_{on} = \frac{n}{n+N_o} \quad ; \quad N_o = \frac{\mathcal{E}v_o(\tilde{\theta})}{\mathcal{V}m_o(\tilde{\theta})} \quad . \tag{4.1}$$

Thus, in the one-dimensional case, $f_o(x_1)$ in (3.8) would be replaced by:

$$f_o(x_1) = m_o + Z_{o1}(x_1-m_o) \quad . \tag{4.2}$$

This means that the exact Bayesian forecast would have the interesting shape shown in Figure 3.

This shape shows us that, if x_1 is small, we believe it is of ordinary type, and we should experience rate according to the linear law (4.2). But as x_1 increases beyond m_o into the region where the odds-likelihood-ratio becomes significant, we begin to hedge our bets on the fact that we have an ordinary observation, and to reduce the dependence of the forecast on x_1. Finally, for x_1

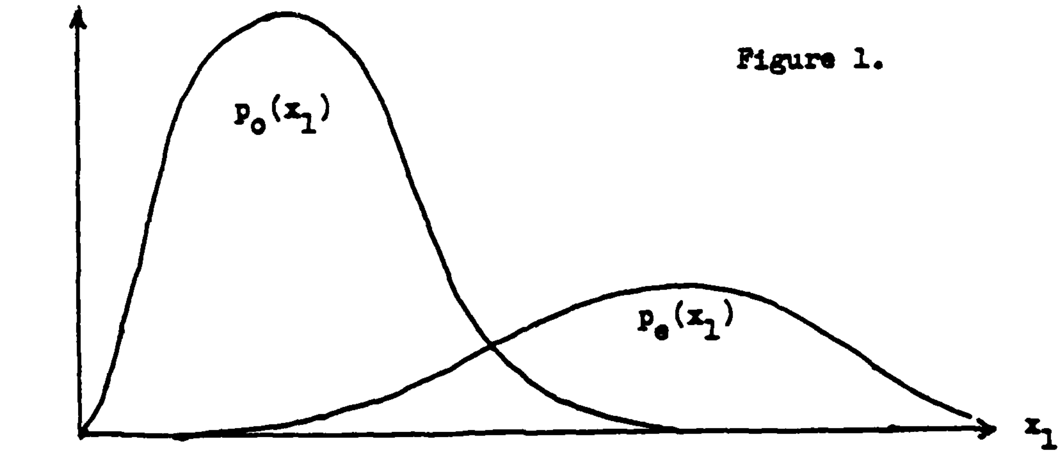


Figure 1.

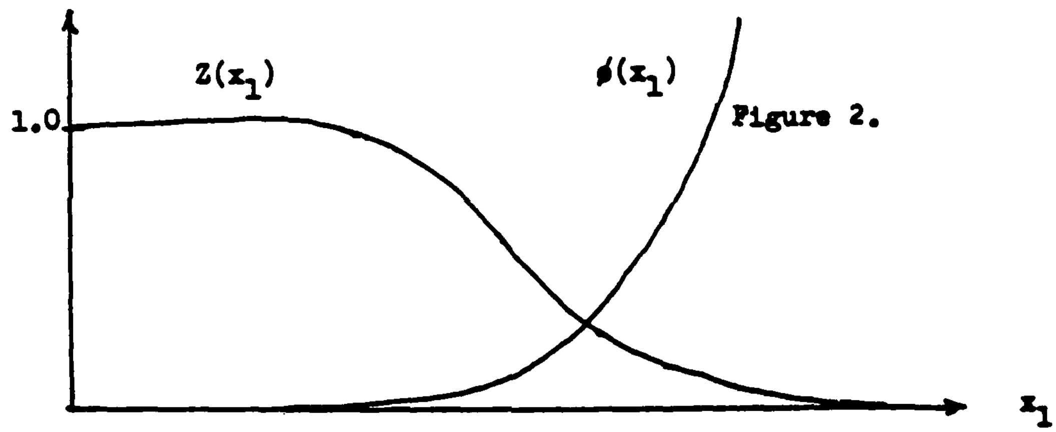


Figure 2.

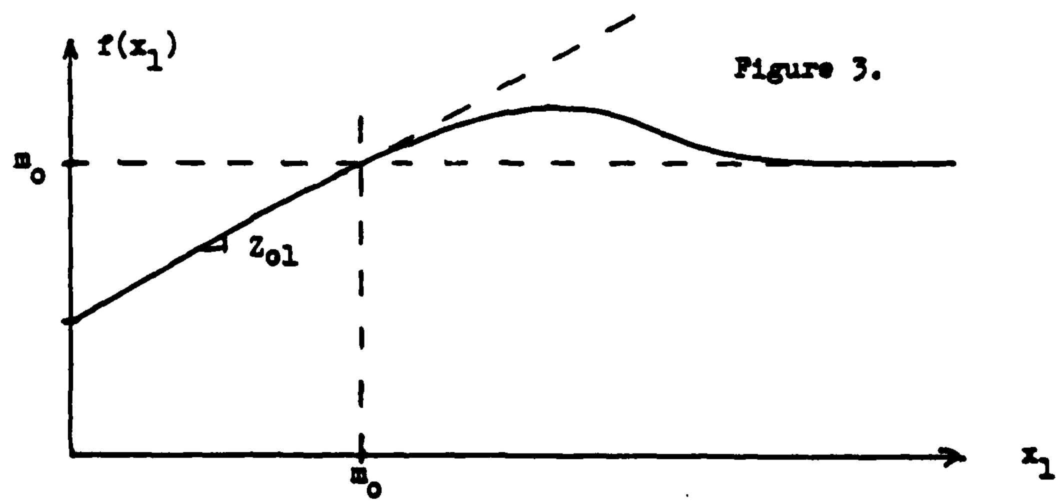


Figure 3.

very large, an excess observation is highly credible, and we settle for the "no-information" manual rating, m_0.

We see that the resulting forecast is quite similar to that obtained by ordinary credibility theory, but trimming the data and replacing x_1 by $\min(x_1,M)$ as in (1.2). Although sharp trimming will not have the "bump" shown in Figure 3, the effect will be small because the three parameters (a,b,M) in (1.2) can be adjusted to minimize the mean-squared error, thus giving a straight-line portion to the forecast which is slightly different than (4.2).

Another point in favor of the trimming is that it might be difficult to implement the exact predictive form in Figure 3 in a real experience-rating scheme; it would be difficult to explain a plan in which a policy with a larger claim might have a smaller next year's premium !

5. A Numerical Example

Figure 4 shows a numerical example in which normal densities were chosen for the average ordinary and excess densities; the means and standard deviations were:

$$m_o = \mu_o = 10 \ ; \ \sigma_o = 5 \ ; \qquad \mu_e = 50 \ ; \ \sigma_e = 20 \ .$$

A contamination probability $\pi = 0.1$ and a credibility factor $Z_{01} = 0.5$ were used, so

$$f_o(x_1) = 10 + 0.5(x_1-10) \ .$$

The resulting exact forecast $f(x_1)$ is plotted in Figure 4, together with the optimal trimmed forecast, which is approximately

$$f(x_1) = 10 + 0.441\left[\min(x_1,14.7) - 10\right] \ .$$

Note that in the use of Gisler's results, one must subtract $\pi\mu_e$ from his forecast, as he does not have an explicit model for the generation of excess claims, and is predicting a future observation of either type.

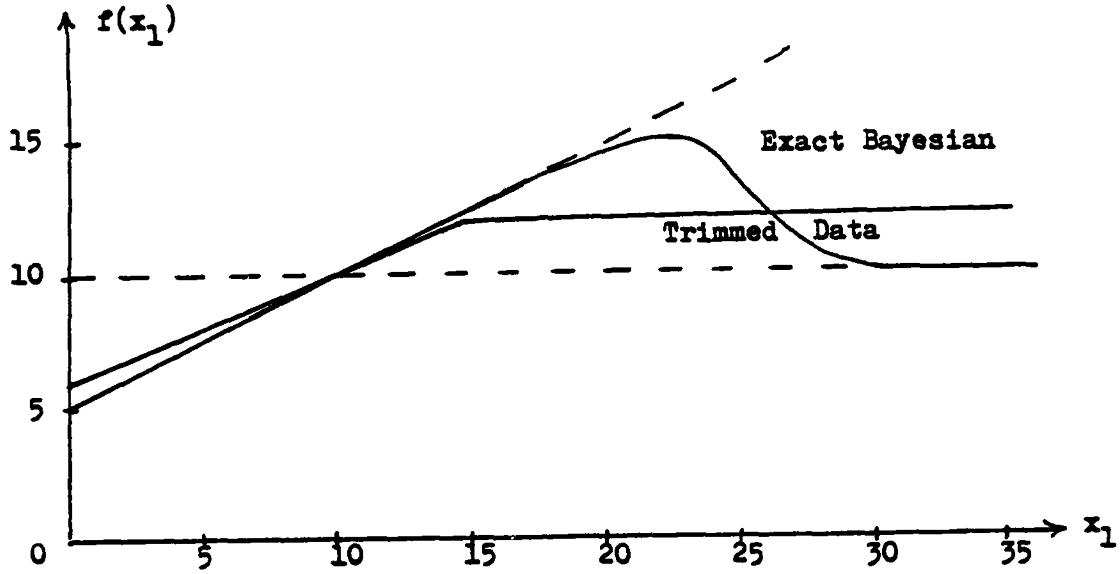


Figure 4. Exact Bayesian and Trimmed Data Forecasts for Example.

6. General Case

From the preceeding, it should be clear that the optimal predictor for an arbitrary number of observations n consists of 2^n terms from (2.2)(2.3), corresponding to all the different ways in which the data $\underline{x}=(x_1,x_2,\ldots x_n)$ can be partitioned into ordinary or excess categories. The formulae are greatly simplified in the general case if we use set-theoretic notation.

Let $\mathcal{N} = \{1,2,\ldots n\}$, $\mathcal{J}$ be any subset of $\mathcal{N}$ (including $\mathcal{N}$ and the empty set $\emptyset$), $\bar{\mathcal{J}} = \mathcal{N} - \mathcal{J}$, and $J = \|\mathcal{J}\|$. Use also $\mathcal{J}$ as a subscript to denote an arbitrary subset of observables, so that, for example, $x_{\mathcal{J}} = \{x_j \mid j \in \mathcal{J}\}$. Then the probability that this subset is all ordinary is:

$$(6.1) \qquad p_0(x_{\mathcal{J}}) = \int \prod_{j \in \mathcal{J}} p_0(x_j \mid \theta)\, p(\theta)\, d\theta \quad ,$$

whereas the probability that it is all excess is:

$$(6.2) \qquad p_e(x_{\mathcal{J}}) = \prod_{j \in \mathcal{J}} p_e(x_j) \quad .$$

For consistency in the following equations, set $p_o(x_\emptyset) = p_e(x_\emptyset) = 1$.

Next, we define $f_o(x_{\mathcal{J}})$ to be the Bayesian forecast of an ordinary random variable, $\tilde{y}_o$, using only the data $x_{\mathcal{J}}$, assumed to be all of ordinary type. This might be the J-term generalization of (4.2), e.g., (2.1) with m replaced by m_o, Z_n replaced by Z_{oJ}, and of course using only the data $x_{\mathcal{J}}$. For consistency, the no-data forecast is $f_o(x_\emptyset) = m_o$.

Then, examination of the expansion of (2.3) shows that the forecast consists of the weighted sum of 2^n forecasts:

$$(6.3) \qquad f(\underline{x}) = \sum_{\mathcal{J}} Z_{\mathcal{J}}(\underline{x}) \; f_o(x_{\mathcal{J}}) \; ,$$

where the data-dependent credibility factors are:

$$(6.4) \qquad Z_{\mathcal{J}}(\underline{x}) = K \; (1-\pi)^J \pi^{n-J} \; p_o(x_{\mathcal{J}}) \; p_e(x_{\bar{\mathcal{J}}}) \; ,$$

and K is adjusted so that the factors sum to unity.

The sum in (6.3) is over all 2^n subsets of $\mathcal{N}$, although, as previously stated, it is unlikely that more than a few excess claims would be present with π small. This suggests the following computational strategy: arrange the data in decreasing magnitude, $x_1 > x_2 > \ldots > x_n$, take $\bar{\mathcal{J}}$ successively to be $\emptyset$, $\{1\}$, $\{1,2\}$, ...etc., and compute the corresponding credibility factors $Z_{\mathcal{J}}$. At some point these factors will become quite small, and the remaining terms in (6.3) can be neglected. One can, if desired, bound the neglected terms.

7. Continuing Research

The results presented here are part of a continuing research effort, joint with H. Bühlmann and A. Gisler. Current effort is devoted to multi-dimensional computations, and comparison of trimmed-data forecasts with the exact Bayesian prediction. Preliminary results indicate that the Gisler approximation continues to be quite good in the multi-dimensional case. These and other results will be be presented in an expanded version of this paper later this year.

References

1. A. Gisler, "Optimum Trimming of Data in the Credibility Model", Bulletin of the Society of Swiss Actuaries, 1980, Heft 3, 313-325 (1980).
2. R. Norberg, "The Credibility Approach to Experience Rating", Scandinavian Actuarial Journal, 1979, No. 4, 181-221 (1979).